Tema 4. DESCRIPCIÓN DE UNA VARIABLE: MEDIDAS DE DISPERSIÓN Y DE FORMA

CONTENIDO:

1. MEDIDAS DE DISPERSIÓN:
   - ✓ Varianza
   - ✓ Desviación típica
   - ✓ Coeficiente de variación
   - ✓ Rango
   - ✓ Recorrido intercuartílico

2. MEDIDAS DE FORMA:
   - ✓ Asimetría
   - ✓ Apuntamiento

Lecturas recomendadas:

•PP. 19-26 de *La Estadística en Cómic*, de L. Gonick y W. Smith.

•Capítulos 4 y 5 de *Introducción a la Estadística para las Ciencias Sociales*, de D. Peña y J. Romo.

EJEMPLO 1:

DOS CONJUNTOS DE DATOS

(Salarios anuales en € de la empresa A)

30700 32500 32900 33800
34100 34500 36000

(Salarios anuales en € de la empresa B)

27500 31600 31700 33800
35300 34000 40600

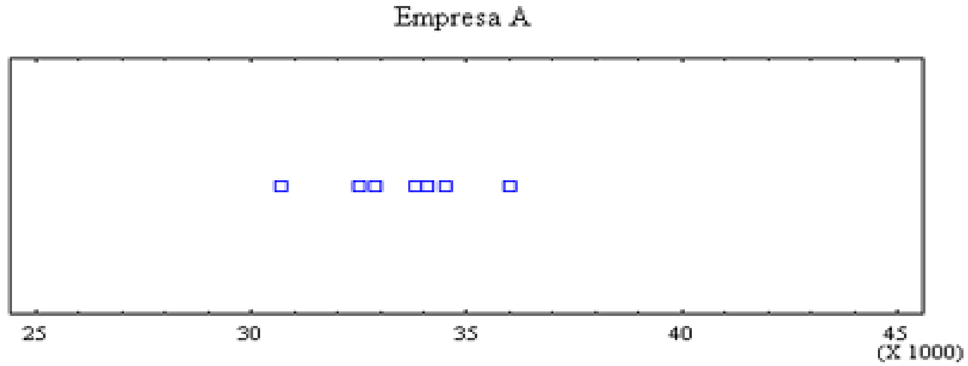
Calcular la MEDIA y MEDIANA de ambos conjuntos de datos:

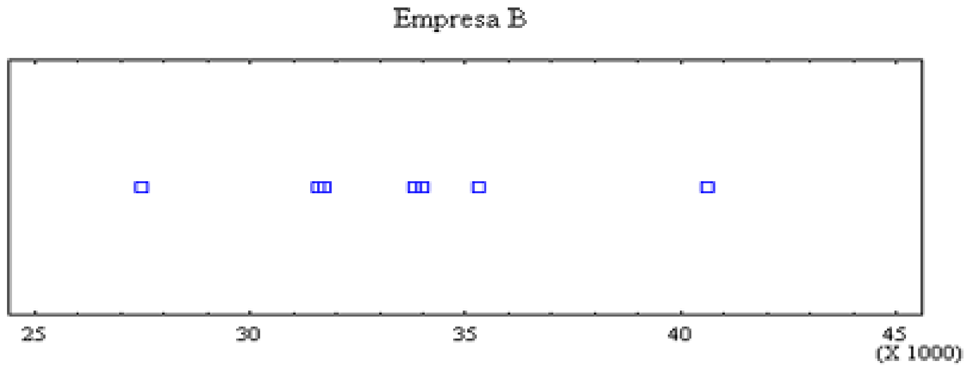
____

____

____

Observa las representaciones gráficas.
Señala la media y la mediana.





Si nos dan solo la Media o la Mediana de unos datos ¿tenemos suficiente información?
¿Cómo podemos medir cuán dispersos están los datos?

EJEMPLO 1: (Continuación)

Calculamos las DISTANCIAS A LA MEDIA:

Empresa A	$x_i - \overline{X}$	Empresa B	$x_i - \overline{X}$
30700	-2800	27500	-6000
32500	-1000	31600	-1900
32900	-600	31700	-1800
33800	300	33800	300
34100	600	34000	500
34500	1000	35300	1800
36000	2500	40600	7100

¿Cuánto suman nuestras dos nuevas columnas?

PROPIEDAD DE LA MEDIA:

$$\boxed{\sum_{i=1}^{N}\left(x_i - \overline{X}\right) = 0}$$

¿Por qué sucede esto?

¿Cómo podemos medir cuán distintos son los datos de su media?

EJEMPLO 1: (Continuación)

Ahora tomando el cuadrado de las distancias a la media:

Empresa A	$(X_i - \overline{X})^2$	Empresa B	$(X_i - \overline{X})^2$
30700	7840000	27500	36000000
32500	1000000	31600	3610000
32900	360000	31700	3240000
33800	90000	33800	90000
34100	360000	34000	3240000
34500	1000000	35300	250000
36000	6250000	40600	50410000
	16900000		96840000

LA VARIANZA:

$$\boxed{\frac{\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2}{N} = S^2}$$

¿Qué mide esta cantidad?
¿Qué unidad tiene esta medida?

DESVIACIÓN TÍPICA: $\boxed{S}$

EJEMPLO 2:

Los importes (en €) de los libros pedidos por los Deptos. de Estadística y de Informática en la última semana son:

Depto. Informática: 329 332 335 323 338

Depto. Estadística: 43 55 32 61 47

Calcula la varianza de los importes para cada depto.

¿En qué depto. han pedido libros con mayor variabilidad de precios?

¿Es justo comparar simplemente la varianza?

COEFICIENTE DE VARIACIÓN

$$\boxed{CV = \frac{S}{|\overline{X}|}}$$

☆ Solo se puede calcular cuando la media sea distinta de "0".

1. Medidas de dispersión

EL RANGO: Es la diferencia entre el mayor y el menor de los datos

EL RECORRIDO INTERCUARTÍLICO:

Es la diferencia entre el tercero y el primero de los cuartiles.

EJERCICIO 1:

Analizamos el volumen de consultas durante el periodo de exámenes en 10 bibliotecas universitarias, y se comparan con las anotadas el año anterior. El % de incremento de consultas fue:

10.2	2.9	3.1	6.8	5.9
7.3	7.0	8.2	3.7	4.3

CALCULA EL RANGO Y EL RANGO INTERCUARTÍLICO

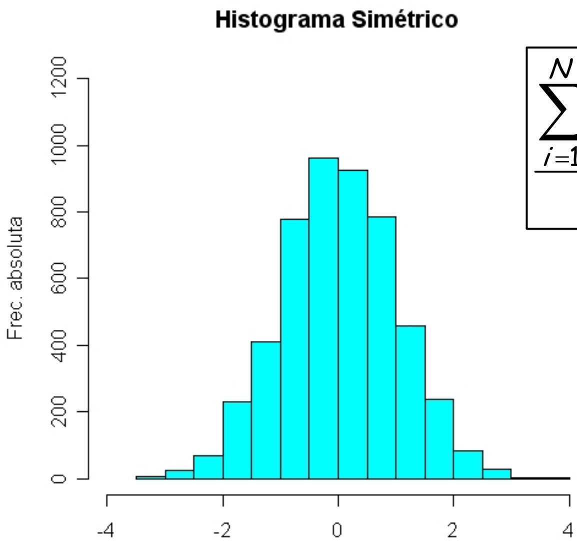
2. Medidas de forma

ASIMETRÍA

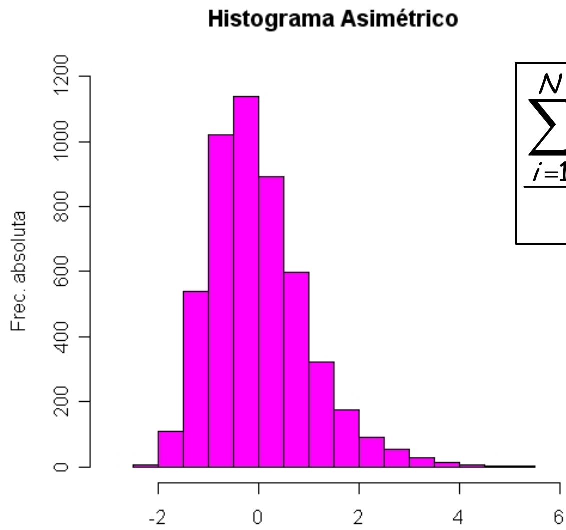
Histogramas de dos conjuntos de datos con media cero y varianza 1:

$$\boxed{\overline{x}=0}$$

$$\boxed{s^2=\frac{\sum_{i=1}^{N}(x_i-\overline{x})^2}{N}=1}$$



$$\boxed{\frac{\sum_{i=1}^{N}(x_i-\overline{x})^3}{N}=0}$$



$$\boxed{\frac{\sum_{i=1}^{N}(x_i-\overline{x})^3}{N}=1}$$

Distribución simétrica: la media deja por la izquierda el mismo nº de observaciones que por la derecha.

Distribución asimétrica por la derecha: Los valores pequeños son más frecuentes que los grandes.

MOMENTO DE ORDEN 3 RESPECTO DE LA MEDIA:

$$\boxed{m_3 = \frac{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^3}{N}}$$

COEFICIENTE DE ASIMETRÍA DE FISHER

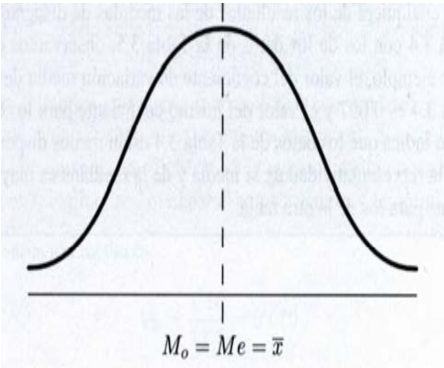
$$\boxed{CA_F = \frac{m_3}{S^3}}$$

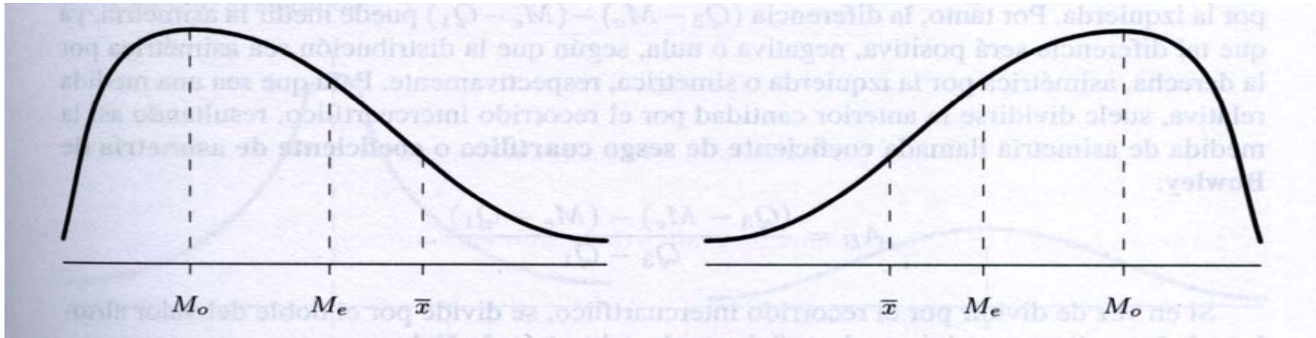
$CA_F=0$ Simétrica

$CA_F>0$ Asimétrica dcha.

$CA_F<0$ Asimétrica izq.

ASIMETRÍA





(a) Distribución asimétrica por la derecha (b) Distribución asimétrica por la izquierda

COEFICIENTE DE ASIMETRÍA DE PEARSON

$$\boxed{CA_p = \frac{\overline{X} - Mo}{S}}$$

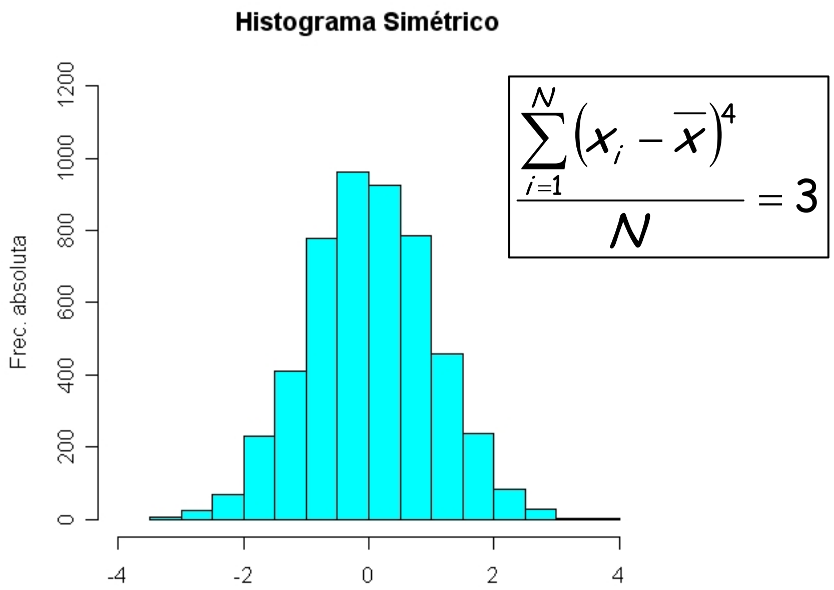
$CA_P=0$	Simétrica
$CA_P>0$	Asimétrica dcha.
$CA_P<0$	Asimétrica izq.

APUNTAMIENTO

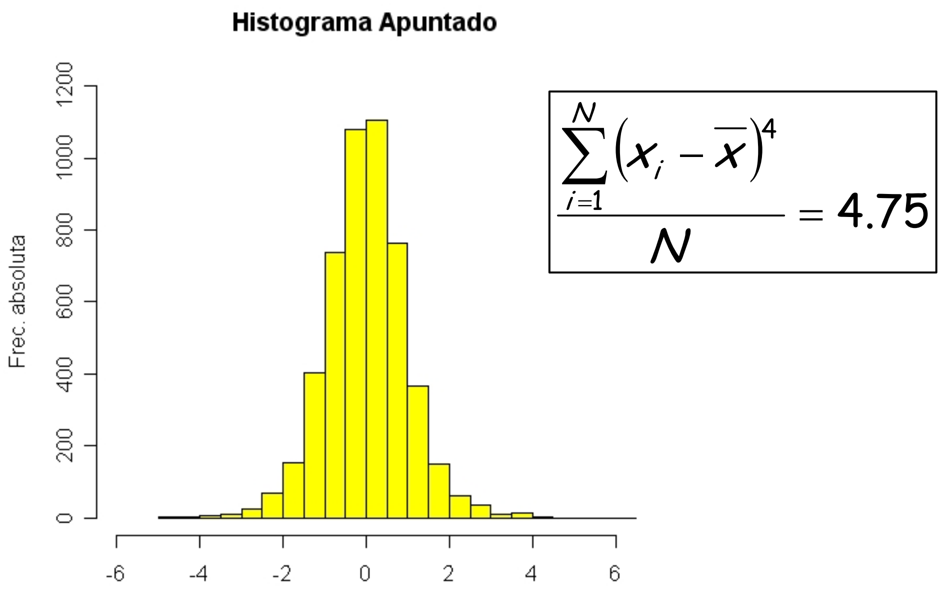
Histogramas de dos conjuntos de datos con media cero y varianza 1: $\boxed{\bar{x}=0}$

$$\boxed{s^2=\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}=1}$$

$$\boxed{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^3}{N}=0}$$



$$\boxed{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^4}{N}=3}$$



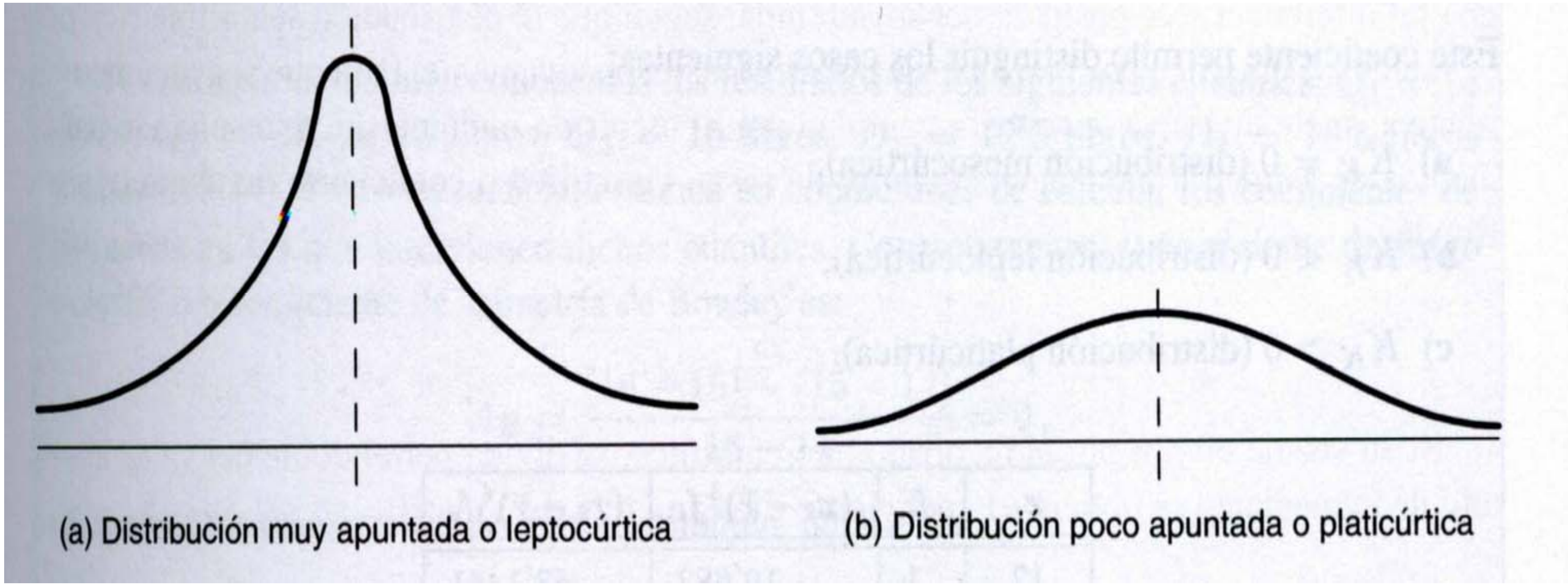
$$\boxed{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^4}{N}=4.75}$$

Distribución no apuntada (mesocúrtica): Los valores centrales son frecuentes.

Distribución apuntada (leptocúrtica): Los valores centrales son muy frecuentes.

2. Medidas de forma

APUNTAMIENTO



COEFICIENTE DE CURTOSIS DE FISHER

$$m_4 = \frac{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^4}{N}$$

$$CC = \frac{m_4}{S^4} - 3$$

CC = 0 (mesocúrtica)

CC > 0 (leptocúrtica)

CC < 0 (platicúrtica)

ACTIVIDADES

ACTIVIDAD 1:

Calcula la media, la varianza, el coeficiente de asimetría de Fisher y el coeficiente de curtosis para las siguientes calificaciones de un grupo de alumn@s:

50	87	63	75	56	63	98	55

ACTIVIDAD 2:

Pon un ejemplo de dos conjuntos de datos con idéntica media pero con varianza muy distinta. Haz un histograma de cada conjunto de datos.

Estandariza cada conjunto de datos: para ello, a cada dato réstale su media y divídelo entre su desviación típica. Vuelve a hacer los histogramas de cada conjunto de datos.

ACTIVIDADES

ACTIVIDAD 3:

¿Puedes explicar por qué el momento de orden 3 respecto de la media, m_3, nos indica la simetría/asimetría de la distribución?

¿Por qué crees que hace falta dividir m_3 entre S^3 para medir la asimetría de una distribución?

ACTIVIDAD 4:

Pon un ejemplo de dos conjuntos de datos con coeficientes de asimetría muy distintos. Haz un histograma de cada conjunto de datos.